Crystal Structure of Human Cystatin D, a Cysteine Peptidase Inhibitor with Restricted Inhibition Profile*[

Received for publication, October 20, 2004, and in revised form, February 10, 2005 Published, JBC Papers in Press, February 23, 2005, DOI 10.1074/jbc.M411914200

Marcia Alvarez-Fernandez‡, Yu-He Liang§, Magnus Abrahamsson‡,§, and Xiao-Dong Su†,

From the ‡Department of Clinical Chemistry, Institute of Laboratory Medicine, Lund University, SE-221 85 Lund, Sweden, §The National Laboratory of Protein Engineering and Plant Genetic Engineering, Peking University, 100871 Beijing, China, and †Department of Molecular Biophysics, Center for Chemistry and Chemical Engineering, Box 124, SE-221 00 Lund, Sweden

Cystatins are natural inhibitors of papain-like (family C1) and legumain-related (family C13) cysteine peptidases. Cystatin D is a type 2 cystatin, a secreted inhibitor found in human saliva and tear fluid. Compared with its homologues, cystatin D presents an unusual inhibition profile with a preferential inhibition cathepsin S > cathepsin H > cathepsin L and no inhibition of cathepsin B or pig legumain. To elucidate the structural reasons for this specificity, we have crystallized recombinant human Arg20-cystatin D and solved its structures at room temperature and at cryo conditions to 2.5- and 1.8-Å resolution, respectively. Human cystatin D presents the typical cystatin fold, with a five-stranded anti-parallel beta-sheet wrapped around a five-turn alpha-helix. The structures reveal differences in the peptidase-interacting regions when compared with other cystatins, providing plausible explanations for the restricted inhibitory specificity of cystatin D for some papain-like peptidases and its lack of reactivity toward legumain-related enzymes.

Cystatins are natural inhibitors of family C1 (papain-like) cysteine peptidases. In mammals, cystatins inhibit peptidases such as cathepsins B, H, K, L, and S both intra- and extracellularly following a reversible, tight binding mechanism (1). The family C1 enzymes are involved in the normal lysosomal turnover of proteins, but they are also implicated in many disease processes, such as tumor invasion and connective tissue destruction on inflammation (2–4).

The cystatins constitute a superfamily of related proteins. The mammalian superfamily members are of three major types (1, 5, 6). Type 1 cystatins (also called stefins) are primarily cytoplasmatic, single-domain proteins composed of ~100 amino acid residues, with no disulfide bridges and no signal peptide. Type 2 cystatins are secreted inhibitors and are also single-domain proteins (but are about 120 residues long) and present two well-conserved disulfide bridges and typical signal peptides. Type 3 cystatins, or kininogens, are multidomain proteins presenting three tandemly repeated type 2 cystatin-like domains.

Chicken egg-white (CEW)\(^1\) cystatin, an avian type 2 cystatin, was the first cysteine peptidase inhibitor for which the three-dimensional structure was determined by x-ray crystallography (7). The structures of two type 1 cystatins have also been determined: human cystatin A (or stefin A), determined by NMR spectroscopy (8) and recently by x-ray crystallography of a complex with cathepsin H (9); and human cystatin B (or stefin B) in complex with papain, determined by x-ray crystallography (10). All three cystatins show the same overall structure, with a five-stranded anti-parallel beta-sheet wrapped around a five-turn alpha-helix. In these cystatin structures the papain-binding site is a tripartite, wedge-shaped edge, formed by the N-terminal segment and the first and second hairpin loops, which are called L1 and L2. There is also a NMR model for the plant inhibitor oryzacystatin, which shows the same “cystatin fold” as the animal cystatins (11). In addition, the structure of a dimeric form of human cystatin C has been published (12). Although this dimeric form of cystatin C is inactive as a papain inhibitor due to shedding of the binding site, each of the two domains formed by three-dimensional subdomain swapping adopt the monomeric cystatin fold.

Despite these quite extensive structural data, detailed knowledge of what determines the specificity profiles of different cystatins is lacking. Cystatin D is a type 2 cystatin thus far found only in human saliva and tear fluid (13). It is produced as a preprotein of 142 amino acid residues, of which the first 20 residues constitute a typical signal peptide (14). Cystatin D was originally found as the product of a gene segment displaying a high degree of homology to the human cystatin C gene (15). Its complete amino acid sequence displays 55% identical residues compared with the cystatin C sequence, with all sequence motifs known to be essential for cysteine peptidase inhibition well conserved (14). However, the inhibition profile...
Crystal Structure of Cystatin D

of cystatin D for human family C1 peptidases is clearly different from that of cystatin C (1, 16). Unlike the latter, cystatin D is unable to inhibit cathepsin B. Besides, it shows a preferential inhibition of cathepsins S over cathepsins H and L (16). By a site-directed mutagenesis approach to alter residues in the N-terminal segments of cystatin D and C, it has been shown that these residues can interact with the non-primed substrate pockets of the enzymes in a substrate-like manner (17). Moreover, evidence was presented that N-terminal sequence differences partly explain the specificity differences between cystatins D and C. However, by analysis of engineered hybrid cystatins, it was apparent that structural differences in the framework of the cystatin molecule also must have a large effect on the inhibitory specificity of cystatin D (17).

It was recently reported that some type 2 cystatins can inhibit mammalian legumain, a cysteine peptidase of family C13, through a novel reactive site located on the side opposite to the papain-binding site (18), in a loop referred to as the back-side loop (BSL). This site results in tight reversible inhibition of pig legumain and is active on human cystatins C, E/M, and F, but not on cystatins A, B, and D. Thus, also with respect to inhibition of family C13 enzymes, cystatin D displays a more restricted and specific inhibition profile than other type 2 cystatins.

In the present study, we have crystallized and determined the three-dimensional structure of recombinant human cystatin D, with the aim of clarifying the structural reasons for its selectivity at target enzyme inhibition.

EXPERIMENTAL PROCEDURES

Expression and Purification of Recombinant Cystatin D—Human Arg10-cystatin D (one of the two allelic variants present in approximately equal proportions in the population) (19) was overexpressed in an Escherichia coli expression system as described previously (13). After expression, the protein was purified by anion exchange chromatography on a Q-Sepharose column (30 × 300 mm2; Amersham Biosciences), followed by size exclusion chromatography on a Superdex 75 10/30 column (Amersham Biosciences) connected to a fast protein liquid chromatography system. The anion exchange chromatography was performed using 20 mM ethanolamine, pH 9.0, containing 1 mM benzamidinium chloride as elution buffer, and the size exclusion chromatography was performed using 50 mM Tris buffer, pH 7.5, with 150 mM NaCl. The fractions of highest purity were pooled and dialyzed against 100 mM Tris buffer, pH 7.5. The protein solution was then concentrated using a Vivaspin column with a cut-off limit of 5000 Da (Vivasience, Lincroft, NJ) to a final concentration of ~8 mg/ml.

Protein concentrations were determined by UV absorption spectroscopy at 280 nm using ε = 18,200 M⁻¹ cm⁻¹ as extinction coefficient ([A]280, 0.1% = 1.29) (17). The purity of the protein in size exclusion chromatography fractions was determined by size- and charge-separating electrophoreses in 16.5% SDS-PAGE gels (20) and 1% agarose gels (21), respectively.

Crystalization—Crystallization plates were prepared using the hanging-drop vapor diffusion method in 24-well VDX plates (Hampton Research, Laguna Nigel, CA). Initial screening of crystallization conditions was done using Crystal Screen kits 1 and 2 (22, 23) (Hampton Research). Five-µl droplets were used in the initial screens (2.5 µl of protein solution and 2.5 µl of precipitant solution), and 6–10-µl droplets were used in optimization trials. Reservoirs contained 12 µl in initial screens and 100 µl at optimization.

X-ray Data Collection and Processing—Room temperature data were collected on a Mar image plate system (Marresearch GmbH, Hamburg, Germany) mounted on a Rigaku RU-200 rotating anode generator operating at 50 kV, 90 mA. Crystals were mounted in a quartz capillary for data collection. A full data set was collected from a single crystal.

Cryo conditions for data collection were worked out using sucrose as cryo coolant. The crystal was equilibrated with the mother liquor in the presence of 15% sucrose for several hours. The crystal was then mounted in a nylone Cryoloop (Hampton Research) and flash-cooled directly in a cold nitrogen stream at about 100 K. Diffraction data were collected at crystallographic beamline BL711 at the MAX-II synchrotron laboratory (Lund, Sweden) using a Mar345 image plate detector (X-ray Research GmbH, Norderstedt, Germany). A typical exposure time was 60 s/frame with 1° oscillation. All data sets were processed using the DENZO and SCALEPACK packages (24).

Structure Determination—The structure of cystatin D was solved by molecular replacement methods. For the room temperature data, this was done by using CEW cystatin (PDB code 1CEW) (7) as a search model. Different modifications, such as poly(A) and poly(S), were tried. The programs AMoRe (25) and Crystallography & NMR System (CNS) (26) were used for the replacement search of the data between 15.0 and 4.0 Å. The molecular replacement solution was refined using the program CNS on the complete room temperature data (30.0 to 2.5 Å). The refined room temperature structure was then used as model for the rigid-body refinement on the cryo data.

Structural Alignment and Graphical Illustrations—Multiple sequence alignment of cystatins with known structures was initially done by the Genetics Computer Group (GCC) Wisconsin Package software. The alignment was modified using the multiple structure alignment obtained with the program Multiple Alignment of Protein Structures (MAPS).2 The structures used in the alignment were obtained from the Protein Data Bank (27): CEW cystatin, PDB code 1CEW (7); dimeric human cystatin C, PDB code 1G96 (12); cystatin A, PDB code 1DVD (8); cystatin B, PDB code 1STF (10); and oryzacystatin, PDB code 1EQK (11). If not otherwise indicated, the amino acid numbering used is that of human cystatin C,2 as previously used for cystatin D and other human type 2 cystatins (19, 28, 29). Graphical representations were prepared with the programs MOLMOL (30) and GRASP (31).

RESULTS AND DISCUSSION

Crystalization of Cystatin D and Cystal Data Collection—Recombinant human cystatin D crystals appeared during the first week in Crystal Screen kit 1 condition 39 (100 mM sodium-HEPES buffer, pH 7.5, with 2% (w/v) polyethylene glycol 400 and 2 M (NH₄)₂SO₄) at 18 °C. Finer grids based on this condition were settled at the same temperature by using either Tris or HEPES as buffer at a pH interval between 6.5 and 8.0 and by varying the ammonium sulfate (0.4–2.4 M) and polyethylene glycol (1–2%, w/v) concentrations. Crystals were obtained under several conditions. They were stable and presented typical shapes as long rods or plates. Two of the well-diffracting crystals were used for structure determination. These crystals were grown at 18 °C in 100 mM Tris, pH 7.5, with 2.4 µM (NH₄)₂SO₄ and 2.5% polyethylene glycol 400.

A room temperature data set was collected from a plate-shaped crystal with dimensions of about 0.5 × 0.3 × 0.1 mm³. The crystal diffracted beyond 2.5 Å, and a full data set could be collected from a single crystal (Table I). A second similar crystal soaked in 15% sucrose was used to collect a data set at about 100 K, giving diffraction beyond 1.8 Å (Table I).

Molecular replacement was used as method to solve the structure of cystatin D from the room temperature data set. This was accomplished using the crystal structure of CEW cystatin as search model. Using AMoRe (25) and CNS with data collected between 15.0 and 4.0 Å, we obtained the same rotational solutions, which were well above background, regardless of which model was used. From the extinction list, two axes were clearly shown as screw axes. Thus, the space groups P2₁2₁2₁ and P2₂₂₂ were both tested for translational search. The space group P2₁2₁2₁ gave the correct solution. The rigid body refinement using CNS (32) with the room temperature data (30.0 to 2.5 Å) lowered the R cryst/Rfree from 0.438/0.438 to 0.349/0.342, respectively. The simulated annealing method was then applied for further refinement. The maps were calculated and inspected, and the residues of the search model were changed to the correct ones. Composite-omit maps were then calculated to remove model bias. A total of 112 residues, from position Ala10 to Val120 (human cystatin C numbering, Figs. 1A and 2), are included in the final room temperature structure.
model. No electron density was detected for the residues in the N-terminal segment before Ala10 (Fig. 1A). This region must thus be disordered in structure. The flexible region around residues 80–84 was difficult to build in before the composite omit maps were made. Thirty-two water molecules were added to the model where strong difference densities (>3σ) were shown and the hydrogen-bond geometry was good. The individual B-factor refinement was applied to the final room temperature model.

The structure solution was straightforward for the cryo data after the room temperature structure was refined. The cryo structure had slight but significant changes in the cell dimensions (Table I). The refined room temperature model was then used for the rigid-body refinement on the cryo data in the subsequent refinement steps by CNS and to add a total of 85 water molecules to the cryo model (Table I).

Despite the significant unit cell changes, particularly on the length of the b axis, the room temperature and cryo structures turned out to be very similar, with root mean square deviation (r.m.s.d.) on the Ca trace of 0.36 Å and an overall r.m.s.d. with side chains of 0.91 Å. At the C1 peptidase binding region, the r.m.s.d. values when comparing the cryo and room temperature structures are 0.23, 0.13 and 0.63 Å for the main chain atoms of the N-terminal part (amino acid residues Gly11-Ala15), the L1 loop (Gln55-Gly59), and the L2 loop (Val104-Asp108), respectively. The relatively large r.m.s.d. value for L2 is attributable to the contribution from Pro105. At the putative legumain (C13 peptidase) binding site, the r.m.s.d. value for the main chain atoms of the BSL (Val37-Glu41) is 0.15 Å. The strong similarity between the room temperature and cryo structures can also be seen from a B-factor plot (supplemental Fig. 1).

Overall Structure—Human cystatin D adopts the so-called cystatin fold (Fig. 1B), as do its five homologues with known structures (7, 8, 10, 11). The core structure is built from a five-stranded anti-parallel β-sheet consisting of β1 (Ile13-Thr16), β2 (Ser44-Ile57), β3 (Val60-Thr71), β4 (Glu95-Val104), and β5 (Lys109-Lys119) that is wrapped around a five-turn α-helix (Lys21-Lys36) (Figs. 1B and 2).

Comparison with CEW cystatin (Fig. 1, C and D) revealed some notable differences in the overall structure of cystatin D: 1) the loop at the C-terminal end of the α-helix is larger than that in CEW cystatin, which deforms the last turn of the helix. 2) Cystatin D does not present a bulge in the middle of the second strand of the β-sheet around position 49. 3) There is no helix in the appendix loop of cystatin D. Instead, it presents a disordered conformation. This is also the case for the monomeric domains in the crystal structure of dimeric human cystatin C (12). 4) Marked differences in the putative peptidase-interacting regions of cystatin D are observed (see below).

Comparison of the electrostatic potential surfaces of the two proteins (Fig. 1E) revealed further differences. Cystatin D has a narrower and more elongated shape than CEW cystatin. This might be a result of the missing bulge in the β2-strand, straightening up the β-sheet in cystatin D. Also, the two cystatins differ quite significantly with respect to the charge distribution on their surfaces. In CEW cystatin, positive and negative charges are evenly distributed on the protein surface. In cystatin D, however, the surface presents some strongly (mainly negatively) charged areas, whereas other areas are pronounced hydrophobic.

Human cystatin D is present in two natural forms due to a gene polymorphism (19). It has been shown that this variation neither significantly affects the enzyme binding properties of the inhibitor nor has drastic effects on protein stability (16), but the structural consequences of the variation have not been elucidated. The form crystallized here, Arg26-cystatin D, in which the 26th residue of the 122-residue predicted mature reference sequence of cystatin C,3 the polymorphic residue is number 25 in an alignment of cystatin sequences (Fig. 2). The strong similarity between the room temperature and cryo structures can also be seen from a B-factor plot (supplemental Fig. 1).

Overall Structure—Human cystatin D adopts the so-called cystatin fold (Fig. 1B), as do its five homologues with known structures (7, 8, 10, 11). The core structure is built from a five-stranded anti-parallel β-sheet consisting of β1 (Ile13-Thr16), β2 (Ser44-Ile57), β3 (Val60-Thr71), β4 (Glu95-Val104), and β5 (Lys109-Lys119) that is wrapped around a five-turn α-helix (Lys21-Lys36) (Figs. 1B and 2).

Comparison with CEW cystatin (Fig. 1, C and D) revealed some notable differences in the overall structure of cystatin D: 1) the loop at the C-terminal end of the α-helix is larger than that in CEW cystatin, which deforms the last turn of the helix. 2) Cystatin D does not present a bulge in the middle of the second strand of the β-sheet around position 49. 3) There is no helix in the appendix loop of cystatin D. Instead, it presents a disordered conformation. This is also the case for the monomeric domains in the crystal structure of dimeric human cystatin C (12). 4) Marked differences in the putative peptidase-interacting regions of cystatin D are observed (see below).

Comparison of the electrostatic potential surfaces of the two proteins (Fig. 1E) revealed further differences. Cystatin D has a narrower and more elongated shape than CEW cystatin. This might be a result of the missing bulge in the β2-strand, straightening up the β-sheet in cystatin D. Also, the two cystatins differ quite significantly with respect to the charge distribution on their surfaces. In CEW cystatin, positive and negative charges are evenly distributed on the protein surface. In cystatin D, however, the surface presents some strongly (mainly negatively) charged areas, whereas other areas are pronounced hydrophobic.

Human cystatin D is present in two natural forms due to a gene polymorphism (19). It has been shown that this variation neither significantly affects the enzyme binding properties of the inhibitor nor has drastic effects on protein stability (16), but the structural consequences of the variation have not been elucidated. The form crystallized here, Arg26-cystatin D, in which the 26th residue of the 122-residue predicted mature cystatin D sequence (13) is Arg, has a population frequency of 0.45 (19). Because the predicted amino acid sequence of mature cystatin D is 1 residue longer in the N terminus than the reference sequence of cystatin C,3 the polymorphic residue is number 25 in an alignment of cystatin sequences (Fig. 2). The other natural form of human cystatin D (population frequency of 0.55) has Cys in this position, unlike other type 2 cystatins,
FIG. 1. Cystatin D structure and comparison with CEW cystatin. A, the amino acid sequence of the cystatin D form crystallized, recombinant human Arg26-cystatin D. The recombinant protein has 2 extra residues (Ala-Pro) in the N terminus but is otherwise identical to one of the two natural forms of cystatin D, with Arg in position 26 of the 122-residue mature protein sequence. The first 12 residues in the N-terminal
for which Arg25 is well conserved (Fig. 2). The present structure shows that the Arg25 residue in cystatin D is situated in the second turn of the α-helix. Although it seems to be exposed at the surface of the protein, its side chain is undoubtedly oriented toward the cavity formed by the bent L2 loop, as revealed by the map. The side chain is likely trapped by a salt bridge with the side chain of either Glu103 or Asp108 in the proximity of the L2 loop. The electron density for the amine groups further out in the side chain of Arg25 is weak or almost absent in both the cryo and the room temperature structures. This indicates a large flexibility in the conformation of these amine groups, reflected by high B-factor values, suggesting that they can alternating between two anchoring sites formed by Glu103 and Asp108. Similarly, the conserved Arg residue in CEW cystatin seems to form a hydrogen bond with the Ser residue in position 103 (Ser115 in CEW cystatin numbering). By homology, the conserved Arg residue in the α-helix of the sequence (33–42). These segments form a tripartite wedge-shaped edge (7, 8, 10, 11, 43) that enters the catalytic site in a substrate-like manner (10, 44). In the cystatin D structure, the papain-binding segments are located in such a tripartite wedge (Fig. 1) in accordance with the other cystatin structures.

The density map for the N-terminal segment in cystatin D is poorly defined, and therefore the structure model gives little information about the potential for interactions between the N-terminal segment of this cystatin and target family C1 peptidases. In the room temperature model, the first residue for which clear electron density is shown is Ala10. In the cryo model for cystatin D, this residue is Gly11. In the structures of CEW cystatin and human cystatin C, the first residues with clear density are Gly11 (Gly7 in CEW cystatin numbering) and Val10, respectively (7, 12). This highlights a similar, very flexible N-terminal segment in all type 2 cystatins. For the other parts of the potential papain-binding site, there are some relevant differences in the cystatin D structure compared with that of CEW cystatin, however.

The L1 loop of the binding site contains the conserved cystatin motif, QXVXG, and is located between strands β2 and β3. L1 in cystatin D is larger and adopts a more “squared” form than it does in CEW cystatin (Fig. 3A). This broader loop is a consequence of the missing β-bulge around position 49 in the β2-strand. This bulge is present in all cystatin structures solved so far, i.e. in CEW cystatin; human cystatins C, A, and B; and the plant cystatin, oryzacystatin. The missing bulge causes a displacement of 1 amino acid residue in the β2-strand of cystatin D, when comparing the sequence-based alignments of previous publications (29) to the three-dimensional structure alignment based on the present results (Fig. 2). Both alignments of the recombinant protein are shown in gray to indicate the poor electron density for this segment. The secondary structure elements are indicated in yellow for α-helix and blue for β-sheet. The motifs known to be important for the inhibition of papain-like enzymes by other cystatins are marked by red boxes, and some well-conserved residues in these are indicated according to human cystatin C numbering. The putative legumain-binding site (BSL) is also indicated, and the presence of an Asn residue in this loop is pointed out (underlined). The position of the residue varying due to a gene polymorphism (Cys/Arg) is marked by an arrowhead. B, ribbon representation of the cryo structure of human cystatin D viewed from the front. The α-helix is marked in yellow, and the β-sheet is marked in blue. The three segments involved in papain binding, formed by the N-terminal segment (N-term) and the first and second hairpin loops (L1 and L2), are indicated. The BSL involved in legumain inhibition by other type 2 cystatins is also indicated. C and D, representations of the aligned structures of human cystatin D (in magenta) and CEW cystatin (in cyan) are viewed from the front and from the C-terminal end of the α-helix, respectively. E, the surface rendering overlapped with structures for charge distribution on CEW cystatin (left) and the cystatin D cryo structure model (right). The color scale on the top shows charge intensity as indicated by the values. The protease binding loops are labeled as described in B. The illustrations were made by the programs MOLMOL and GRASP.
should interact with the wider part of the cleft, where the conserved Trp^{106} in cystatin D may interact with the imidazole rings of Trp^{177} and Trp^{181} in papain, as most likely is the case for the corresponding Trp residue in other type 2 cystatins.

Cystatin D and papain should fit well together from the electrostatic point of view because the very hydrophobic active site cleft of papain and the rather hydrophobic wedge of cystatin D should complement each other without any substantial hindrance. However, sterical hindrances caused by the local topology of cystatin D are likely the main factor weakening the binding. If cystatin D would fit into the active site groove in the same way as cystatin B does, in order to achieve the largest surface of contact with the enzyme, we would expect the side chain of Val^{57a} in the inhibitor to collide with the walls of the cleft, most likely with Trp^{177}. This means that cystatin D might not be able to enter the enzyme cleft as deeply as other cystatins probably do. This results in lost contacts between the other cystatin parts involved in enzyme binding and the peptidase and, consequently, decreases the papain affinity of cystatin D compared with CEW cystatin and human cystatin C (1). Moreover, even if L1 did not protrude as much as it does, the more bent L2 might not be able to make as many contacts with papain as expected by analogy to the mode of enzyme binding of other cystatins.

Previous studies have shown that a truncated form of human cystatin C lacking the first 10 residues of the N-terminal segment has 3 orders of magnitude lower affinity for papain and other family C1 peptidases than full-length cystatin C (34, 38). The same large decrease in target enzyme affinity is observed for its W106G variant (39). The affinity of cystatin D for papain ($K_i$ 1.9 nx; reviewed in Ref. 1) is 5 orders of magnitude lower than that observed for wild-type cystatin C. These kinetic data suggest that sterical hindrances due to the larger L1 loop of cystatin D destabilize contacts not only in this region but also in the N-terminal region and/or in the L2 loop. In addition, the bent L2 loop most likely disfavors interactions between the imidazole rings in the inhibitor (Trp^{106}) and the enzyme (Trp^{177} and Trp^{181}).

Likewise, we docked the structures of cystatin D and cathepsin B (PDB code 1HUC, Ref. 52), again using the complex between cystatin B and papain as starting point for the model. Contrary to the inhibition of papain by cystatins, cathepsin B is inhibited in a two-step kinetic reaction (45). As proposed by Nylander et al. (45), the first step is regulated by the anchoring of the N-terminal part of the cystatin in the non-primed S pockets. This is followed by displacement of the occluding loop of cathepsin B as the anchored cystatin pushes it away in order to introduce the L1 and L2 loops in the S' subsites. Structural studies have shown that the N-terminal segment of cystatin D is less favorable for this initial interaction than the N-terminal segment of cystatin C (39). In effect, when the N-terminal segment of cystatin D (the region until Gly^{59}) was added on to a cystatin C framework in a hybrid cystatin molecule, this variant showed 30 times lower affinity for cathepsin B than wild-type cystatin C. Still, it inhibited the enzyme (39). On the other hand, the introduction of the N-terminal segment of cystatin C into cystatin D in another hybrid molecule did not alter the inability of wild-type cystatin D to inhibit cathepsin B (39). The latter result is most likely due to the fact that cystatin D, even if equipped with the more effective N-terminal segment of cystatin C, fails to push away the occluding loop and bind to the peptidase with the other two segments of the inhibitory wedge (the L1 and L2 loops). This is likely due to the side chain of Val^{57a} in the inhibitor being located too close to Trp^{221} in the S'_1 pocket of cathepsin B. Furthermore, the L2 loop of cystatin D is marked by the presence of two negatively charged groups,
i.e. in Glu107 and Asp108. These negatively charged side chains would be situated in an unfavorable electrostatic environment established by the also negatively charged Asp224 in the active site cleft of cathepsin B, if cystatin D was forced to interact with the enzyme in the same way as cystatin B (stefin B) does with papain (10). In the same positions, CEW cystatin and human cystatin C present non-charged residues. In the case of cystatin B, although it presents a Glu residue in the loop, its positively charged His106 (His104 with CEW cystatin numbering) should fit well into the pocket.

The Putative Binding Site for Legumain-like Peptidases—As recently reported, some type 2 cystatins are able to inhibit mammalian legumain (18), a lysosomal cysteine endopeptidase of family C13 (46), which shows preference for hydrolysis after an asparagine residue (47). The BSL at the end of the main α-helix (Fig. 1, A and B), containing residue Asn39, was identified as most likely being directly involved in legumain inhibition by these cystatins. In effect, Asn39 could be responsible for an inhibitory mechanism in which cystatins inhibit mammalian legumain in a substrate-like manner (18). Like the inhibitorily active cystatins identified so far (human cystatins C, E/M, and F; CEW cystatin; and Bm-CPI-2, a type 2 cystatin homologue from the filarial nematode parasite Brugia malayi (GenBankTM accession number AF015263) (53)), cystatin D presents an asparagine residue in this loop (Figs. 1, 2, and 4).

Even so, cystatin D is the only human type 2 cystatin investigated that cannot inhibit mammalian legumain (18). Assuming that Asn39 in other type 2 cystatins indeed is directly involved in legumain inhibition, we examined and compared the structures in the BSLs of cystatin D and CEW cystatin (Fig. 4), the latter of which is a tight binding inhibitor of pig legumain (48). This was done in an attempt to provide a plausible explanation of why cystatin D is inactive as a legumain inhibitor.

For cystatin D, the structural differences observed were as follows: 1) the Asn residue in cystatin D is neither structurally conserved nor accessible at the surface of the protein. Instead, it is located at a position corresponding to residue 38 in the other type 2 cystatin structures (residue 36 in CEW cystatin) (Fig. 4A). Its side chain points toward the core of the protein and is strongly hydrogen-bonded to Lys75 at the end of the third β-strand. 2) Structurally, there is an insertion of 1 amino acid residue in the loop. The isoleucine residue located between Val37 and Asn38 deviates the most from the Ala37 and Ser38 residues in the corresponding loop segment of CEW cystatin (Ala35 and Ser36 in CEW cystatin numbering). This Ile residue is surrounded by the hydrophobic environment provided by the side chains in the C-terminal end of the α-helix and those in the end of the fifth β-strand. As a consequence of this stabilization, the Ile residue deforms the loop and buries Asn38 deeper into the core. 3) Cystatin D presents a lysine residue at the position corresponding to residue 39 (residue 37 in CEW cystatin), instead of the conserved asparaginyl residue believed to be involved in legumain inhibition in other type 2 cystatins (18). This lysine residue is surrounded by the solvent and hence is accessible on the protein surface (Fig. 4, A and B). Legumain activity is specific for the hydrolysis of substrates with an asparaginyl residue in the P1 position (47), showing preference for Asn residues located in hydrophilic surface loops. Although accessible for enzyme binding, Lys39 is far from being adequate bait for the S1 pocket in the enzyme. Thus, there is little reason to believe that legumain would show any affinity for the BSL in cystatin D.

One interesting possibility indicated by the finding that Lys39 is exposed in the BSL of cystatin D and situated as Asn39 in cystatin C is that cystatin D may have evolved as an inhibitor of enzymes with a preference for Lys binding in their S1 pockets. The inhibition of legumain by some cystatins proves that clan CD enzymes with overall similarity to legumain of family C13 could generally interact well with the BSL binding region. Good candidates for cystatin D target enzymes could e.g. be the Lys-gingipains of family C25 from the periodontal bacterium Porphyromonas gingivalis, which have a strict preference for Lys bonds in substrate polypeptides (49, 50). It is tempting to speculate that cystatin D acts as a physiological inhibitor of these or similar enzymes in saliva and hence could have a biological function to inhibit the growth and action of pathogenic oral bacteria.

CONCLUSIONS

In the present study, we have determined the structure of recombinant human cystatin D by x-ray crystallography under both room temperature and cryo conditions (2.5 and 1.8 Å resolution, respectively). This type 2 cystatin is not as widely distributed in the body as its homologue, cystatin C, but rather is restricted to saliva and tear fluid (13). This may point to a more restricted biological function of cystatin D than that of a general protector against papain-like lysosomal peptidases being released from tumor cells or leaking from dying cells, as has
been suggested for cystatin C. Besides, whereas cystatin C is considered a universal inhibitor, displaying inhibitory activity against all family C1 peptidases studied without relevant specificity, cystatin D shows a much more restricted inhibition profile with affinity for cathepsin S > cathepsin H > cathepsin L and no inhibition of cathepsin B or pig legumain in family C13 (16, 18). This restricted inhibition profile makes cystatin D a good target for structure-function studies aimed at an understanding of factors determining the inhibitory specificity of cystatins.

The crystal structures of cystatin D reveal no exceptional overall differences between this cystatin and its homologues. The cystatin fold is rather well conserved, leaving the major structural differences to the most flexible parts of the protein, i.e., the peptidase-binding sites. Radical differences in the topology of the L1 loop in cystatin D, containing the conserved cystatin motif involved in C1 peptidase inhibition, are likely the major reason for the restricted inhibition profile of cystatin D with respect to interaction with family C1 peptidases. The larger L1 loop in cystatin D might indicate an ability for a deeper and more selective interaction with a specific target enzyme than we would expect for the more general inhibitor cystatin C.

Structural differences in the putative binding site for family C13 peptidases are clearly present in cystatin D and are most likely the reason why cystatin D is not an inhibitor of mammalian legumain. An equivalent to the Asn39 residue present in type 2 cystatins with ability to inhibit legumain is not present in the cystatin D structure. The Asn residue positioned in a nearby location in the cystatin D back-side loop is not accessible at the surface but rather buried in the interior of the structure. This indirectly supports a model for legumain inhibition by cystatins that relies on a substrate-like interaction between the Asn39 residue and the $S_1$ pocket of the active site cleft of the enzyme.

Acknowledgments—We thank Drs. Matthias Bochtler and Maria Häkansson for helpful comments and fruitful discussions.

REFERENCES

Crystal Structure of Human Cystatin D, a Cysteine Peptidase Inhibitor with Restricted Inhibition Profile
Marcia Alvarez-Fernandez, Yu-He Liang, Magnus Abrahamson and Xiao-Dong Su

doi: 10.1074/jbc.M411914200 originally published online February 23, 2005

Access the most updated version of this article at doi: 10.1074/jbc.M411914200

Alerts:
• When this article is cited
• When a correction for this article is posted

Click here to choose from all of JBC's e-mail alerts

Supplemental material:
http://www.jbc.org/content/suppl/2005/03/16/M411914200.DC1

This article cites 52 references, 22 of which can be accessed free at
http://www.jbc.org/content/280/18/18221.full.html#ref-list-1